



Signaler “ à la volée ” les ruptures et les frontières prosodiques : une technique comportementale

Monique Vion, Annie Colas

► To cite this version:

Monique Vion, Annie Colas. Signaler “ à la volée ” les ruptures et les frontières prosodiques : une technique comportementale. Revue PAROLE, 2007, 41/42, pp.79-118. hal-00308394

HAL Id: hal-00308394

<https://hal.science/hal-00308394>

Submitted on 30 Jul 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**SIGNALER « A LA VOLEE »
LES RUPTURES ET LES FRONTIERES PROSODIQUES :
une technique comportementale**

M^{me} nique Vi^{en} & Annie C^{las}

Laboratoire Parole et Langage (CNRS-UMR 6057)
Université de Provence
29 avenue Robert Schuman
F-13621 Aix-en-Provence

[m^{me} nique.vi^{en}@univ-pr^ovence.fr](mailto:m`nique.vi`n@univ-pr`vence.fr)

Résumé

L'article présente un paradigme en temps réel dont le but est d'étudier la perception des ruptures et des frontières de discours chez des experts et des auditeurs naïfs.

On compare la perception des ruptures et des frontières prosodiques par des auditeurs naïfs lors de l'écoute de quatre versions d'un échantillon d'enregistrement de parole spontanée (original, filtré, resynthétisé et resynthétisé avec conversion de phonèmes). Les résultats obtenus en temps réel sont ensuite confrontés à l'analyse de la version originale de l'échantillon (type de contours d'intonation, niveau de jointure entre les mots) par deux experts.

D'une version à l'autre, le nombre de ruptures localisées par 40% des participants est similaire à celui des principales ruptures intonatives signalées par les experts. Les ruptures sont principalement localisées à une frontière intonative ou sur un élément phatique. Les auditeurs naïfs les indiquent avant la pause silencieuse subséquente quand le signal présente une montée ample et rapide de la F0 dans la partie haute du registre du locuteur ou bien lorsque la hauteur est maintenue constante dans la partie basse du registre.

Abstract

This paper presents an on-line paradigm aimed at studying the perception of breaks and discourse boundaries by both expert and naive listeners.

We compared the perception of breaks and discourse boundaries by naive listener across four versions of a spontaneous speech sample (original, low-pass, filtered, re-synthesized with phoneme conversion). Then, naive listeners' perception of breaks and boundaries were compared with two different experts' off-line analyzes of the original sample (type of intonation contours, degree of separation between words).

Across all versions the number of breaks located by 40% of the participants was similar to the major intonation breaks signaled by the experts. Breaks were mainly located at an intonation boundary or in a phatic portion of the signal. Naive listeners signaled them before the following silent pause when the signal offered a rapid and wide rising pitch in the higher part of the speaker's range or a steady pitch in the lower part of the speaker's one.

Titre courant : localiser les ruptures prosodiques en temps réel

Running head: an on-line technique to locate prosodic breaks and boundaries

Mots clés : parole spontanée, prosodie, mesure en temps réel

Key words: spontaneous speech, prosody, on-line technique

Qui ne s'est pas trouvé un jour, dans le métro ou dans une salle d'attente de gare ou d'aérogare, témoin d'un échange de paroles qui pouvait être vif, enjoué ou profondément grave et pourtant complètement impénétrable. Comment les personnes présentes pouvaient-elles se comprendre au moyen de ce flot de bruits et de sons ? Etre témoin d'un échange verbal dans sa propre langue ne pose en général pas de problèmes majeurs de compréhension d'un flot sonore dans lequel les spécialistes savent pourtant que des indices acoustiques simples comme les silences n'indiquent pas forcément la limite d'unités linguistiques. Dans la vie courante, différents types d'information relevant de l'organisation prosodique, phonologique, lexicale, syntaxique, du contenu sémantique et de la situation d'énonciation, œuvrent à la segmentation du langage parlé et du discours. Et pour les spécialistes, de nombreuses questions se posent à leur propos, comme celle du rôle spécifique des indices prosodiques qui nous intéresse ici.

Dans le champ de la linguistique, les recherches qui abordent la question du rôle de la prosodie dans la segmentation du flux continu de la parole et dans l'interprétation de la structure du discours font le plus souvent appel à des techniques d'écoute réitérée permettant l'annotation de corpus transcrits orthographiquement sans ponctuation. Plus rares sont les recherches qui exploitent des dispositifs de segmentation prosodique "à la volée" lors d'une écoute unique (Lehiste, 1979 ; Swerts & Geluykens 1993, 1994). Le but de l'analyse prosodique subjective est de localiser dans le signal de parole des ruptures¹ et/ou des frontières prosodiques². Ce repérage permet d'analyser ensuite les caractéristiques acoustiques de chacune ainsi que celles des groupements de mots ainsi délimités (Grosz & Hirschberg, 1992; Hirschberg & Nakatani, 1996, Nakatani, Hirschberg & Grosz, 1995 ; Strangert & Helner, 1995 ; Swerts, 1997; Portes 2000 et 2002 ; Auran, Portes, Rami & Rigaud, 2001 ; Strangert, 2004). Il est malaisé d'approcher par ces études le rôle effectif joué par les différentes caractéristiques acoustiques décrites, comme par exemple connaître la contribution des pauses silencieuses à la perception des frontières (Cho & Hirst, 2006).

Dans le champ de la psycholinguistique où la prosodie est maintenant reconnue comme l'une des sources d'information importantes pour le traitement du langage parlé (Frazier, Clifton & Carlson, 2004; Frazier, Carlson, & Clifton, 2006), un corps de connaissances s'est progressivement constitué issu d'études sur son rôle lors de la reconnaissance des mots (traitement pré-lexical et lexical), de la détermination de la structure syntaxique et du repérage de l'organisation du discours (voir les synthèses de Cutler, Dahan, & Donselaar, 1997 ; Fon, 2002 ; Di Cristo, 2004). L'analyse du déroulement temporel de l'activité cérébrale apporte depuis peu la preuve de la participation des indices prosodiques au traitement en temps réel de la parole continue (Astesano, Besson, & Alter, 2004 ; Hruska, Alter, Steinhauer, & Steube, 2001 ; Sanders & Neville, 2003 ; Steinhauer, Alter, & Fiederici, 1999 ; Steinhauer, & Friederichi, 2001, Steinhauer, 2003).

¹ Rupture ou Prosodic Break : « perceptively relevant variation in the speech continuum such as to cause the parsing of the continuum into discrete prosodic units ». Danieli, Garrido, Moneglia, Paniza, Quazza, & Swerts, (2004), p 1511.

² Les ruptures formellement introduites dans le discours par la génération des unités intonatives constituent des frontières prosodiques, à savoir des marques fonctionnelles qui indiquent l'achèvement (momentané ou définitif) à la fois d'un événement acoustique et de ce que le locuteur tient à dire.

Les nombreuses études comportementales qui, depuis une vingtaine d'années, participent à l'établissement de ce corps de connaissances ont cependant des champs d'investigation limités eu égard à la complexité de l'échange de paroles. Les séquences de parole qui obéissent aux contraintes de la méthodologie expérimentale sont le plus souvent des énoncés isolés, voire des portions d'énoncés, souvent issus de la lecture oralisée. Pour contrôler expérimentalement les indices prosodiques (parfois limité à un seul indice), les tâches (détection de cible, poursuite en écho, dévoilement graduel, jugement en choix forcé, etc.) exploitent des séquences de langage parlé bien éloignées des corpus de paroles authentiques. A l'heure actuelle, le rôle des indices prosodiques dans la compréhension de la structure du discours et sa représentation en mémoire, de même que l'incidence de l'activité conceptuelle et de son rythme sur la réalisation prosodique et textuelle, restent des questions empiriquement peu abordées (Cohen, Douaire & Elsabbagh, 2001 ; Vion & Colas, 2005, soumis).

L'analyse auditive subjective et les tâches comportementales sollicitent une activité consciente de la part de l'auditeur, tant au niveau de l'écoute que du mode de réponse. Cette activité ne se limite pas à la compréhension tacite des énoncés entendus (le décodage n'y est pas directement observable, Esser, 2000). L'activité réalisée est dirigée par les jugements que l'auditeur effectue (identification, discrimination, catégorisation, jugements de similarité, d'incongruité, etc.).

Pour résumer, dans le champ de l'étude de la valeur fonctionnelle de la prosodie, l'emploi de tâches comportementales peut induire, de par leur caractère fortement métalinguistique, diverses stratégies de traitement des stimulus, étrangères au processus ordinaire de décodage et de ce fait biaiser l'interprétation des observations.

L'idée d'exploiter un dispositif de réponse pour évaluer l'impact des indices prosodiques au cours même de l'audition des énoncés - objectif de la présente étude - n'est pas neuve. Le Hirste (1979) par exemple a employé un dispositif de ce genre pour étudier les propriétés des frontières de phrases et de paragraphes. Il s'agissait pour les auditeurs d'appuyer en cours d'audition, une ou deux fois consécutives sur une clé de télégraphe selon le cas identifié. L'emploi d'un tel dispositif n'exempte pas la tâche de tout caractère métalinguistique. Il repose sur l'hypothèse qu'un auditeur tout-venant peut exploiter de façon explicite et consciente, sa capacité spontanée à repérer en temps réel le regroupement des mots grâce au prélèvement d'indices prosodiques. Pour préciser l'hypothèse, il semble vraisemblable de soutenir que lorsqu'il emploie un tel dispositif en vue de la segmentation des unités prosodiques l'auditeur ne signale en ligne que les variations (ruptures et/ou frontières) qui sont perceptivement pour lui les plus fortes.

Le dispositif présenté ici trouve ses racines dans l'intérêt porté en psychologie cognitive à l'approche des processus de traitement en temps réel et au rôle des indices prosodiques dans la compréhension du langage parlé. Facilement utilisable aussi bien par un auditeur expert que tout-venant, il permet le traitement d'échantillons variés (parole fabriquée pour les besoins d'une recherche, texte oralisé ou parole spontanée, monologue ou échange de paroles) au moyen de deux outils. L'outil de segmentation permet à l'auditeur de signaler au fur et à mesure de l'acquisition du flux de parole, des phénomènes prosodiques précisés par le chercheur et pouvant être ou ne pas être explicitement définis auprès de l'auditeur. Bien que reposant sur les capacités métalinguistiques de l'auditeur, l'outil de segmentation permet de signaler les phénomènes objets de l'étude, dans une situation qui respecte la forme d'acquisition de l'information linguistique propre aux conditions ordinaires de traitement du langage parlé. Plus classiquement, le dispositif permet à l'auditeur de réexaminer dans un second temps le résultat de sa segmentation. Ce second outil lui permet d'évaluer pour chaque segment un certain

nombre de propriétés définies au préalable et d'exprimer des jugements à leur égard au moyen d'une échelle de réponses à échelons variables (Auran, Colas, Portes & Vion, 2005)³.

L'expérience qui suit est destinée à tester l'outil de segmentation. Elle compare le signalement des variations prosodiques perçues par des auditeurs tout-venant, dans un échantillon de discours authentique présenté sous quatre versions différentes (originale, filtrée, re-synthétisée et re-synthétisée avec conversion de phonèmes). Avec pour hypothèse que les indices prélevés par les auditeurs dépendent du niveau de traitement psycholinguistique que la version entendue leur permet d'atteindre.

METHODE

Tâche

La tâche consiste à segmenter lors d'une unique écoute, l'enregistrement d'une suite de paroles jamais entendue auparavant. L'auditeur est invité à appuyer sur une touche du clavier d'un ordinateur dès qu'il lui semble que, en se fondant sur la prosodie, un mot ou un groupe de mots « forme(nt) un tout ». Au préalable, l'auditeur est invité à utiliser le dispositif de réponse pour segmenter de brefs énoncés extraits du même corpus.

Matériel

Quatre versions d'un même échantillon de discours ont été soumises à l'analyse perceptive de quatre groupes de participants. Les énoncés destinés à la familiarisation avec le dispositif, ainsi que l'échantillon expérimental sont extraits du corpus d'un locuteur masculin enregistré lors d'une émission radiophonique. Il s'agit d'un corpus intégralement transcrit, annoté et analysé par Portes (2004).

Deux énoncés d'environ 10 secondes ont été sélectionnés pour la phase de familiarisation. Le premier est extrait d'un moment de l'enregistrement où le débit du locuteur est relativement lent. Le second est extrait d'un moment de l'enregistrement où son débit est relativement plus rapide.

L'échantillon expérimental est une séquence de paroles d'environ une minute. En adoptant la typologie des dix intonations de base du Français proposée par Delattre (1966)⁴, Portes (2004) a identifié dans cette portion de discours 41 contours intonatifs. La version originale de l'échantillon ainsi que l'annotation de l'intonation sont présentés en annexe (lignes a et b).

La version filtrée de l'échantillon expérimental a été obtenue par l'emploi d'un filtre passe bas de 400Hz. Elle a été réalisée au moyen du logiciel Praat (Boersma & Weenink 2005), par

³ La réalisation du dispositif et son test sont devenus possibles grâce au travail engagé dans le groupe interdisciplinaire « Prosodie et discours », transversal aux équipes du Laboratoire Parole et Langage qu'a animé Albert Di Cristo. Cristel Portes a mis à la disposition du groupe l'échantillon de discours qui a servi à tester l'outil de segmentation ainsi que son analyse en contours intonatifs. Cyril Auran a réalisé l'analyse acoustique automatique de l'échantillon selon divers paramètres prosodiques. Cartherine Chanut, Roxane Bertrand et Albert Di Cristo ont pour leur part évalué les niveaux de jointure entre les mots de ce même échantillon. Enfin et surtout, Cyril Auran a conçu les différents programmes de gestion des outils de segmentation et d'évaluation dont nous avons le projet.

⁴ Les éléments du système intonatif du Français sont conçus par Delattre (1966) et à sa suite par Rossi (1999) comme des morphèmes intonatifs ; chaque contour intonatif étant défini comme un mouvement contrastif de la fréquence fondamentale en relation avec une fonction spécifique.

l'application de la commande « Filter... (formula) » et de la formule : « if x>400 then 0 else self fi ; ».

La version re-synthétisée a été obtenue au moyen du logiciel MBROLA (Dutoit, Pagel, Pierret, Bataille & van der Vreken, 1996)⁵.

La version avec conversion de phonèmes (appelée désormais : version « saltra~dzajn ») a été également obtenue au moyen du logiciel MBROLA, après remplacement des phonèmes originaux par d'autres phonèmes selon une correspondance (Tableau 1) qui s'inspire de la table de conversion employée par Ramus, Hauser, Miller, Morris, & Mehler (2000).

#Tableau 1: table de correspondance des phonèmes pour la version saltra~dzajn (transcription SAMPA)

Phonèmes originaux	Remplacés par
Voyelles orales (i e E a A O o u y 2 9 @)	a
Voyelles nasales (e~a~o~9)	a~
Semi-consonnes (j w H)	j
Plosives sourdes (p t k)	t
Plosives voisées (b d g)	d
Fricatives sourdes (f s S)	s
Fricatives sonores (v z Z)	z
Liquide (l)	l
Liquide (R)	R
Nasales (m n N)	n

Pr`cédure

L'expérience, pilotée par ordinateur (un Apple iBook G4, système os X 10.3)⁶, s'est déroulée en chambre sourde. Les participants, munis d'écouteurs, ont effectué la tâche individuellement. L'expérimentateur présente tout d'abord l'objectif général dans lequel s'inscrit la recherche (l'étude des indices prosodiques sur lesquels les auditeurs se fondent pour structurer et comprendre le langage parlé) ainsi que le principe de la tâche. Celle-ci consiste à écouter l'enregistrement d'une séquence de langage parlé et, en se fondant sur la « musique » de la parole, à la découper en blocs au fur et à mesure de son audition. Un bloc était explicitement défini à l'instar de Strangert (2004) comme un mot ou un groupe de mots qui semble(nt) former un tout. L'expérience se déroule ensuite de la façon qui suit.

Tout d'abord, l'expérimentateur invite l'auditeur à placer les écouteurs et lui soumet quatre brèves séquences de paroles énoncées par un locuteur féminin⁷. Ces séquences sont destinées à focaliser son attention sur le groupement des mots dans les énoncés réalisés par la prosodie. Pour chaque séquence, l'expérimentateur demande à l'auditeur de dire immédiatement après l'écoute, combien de blocs comporte la séquence.

Ensuite, l'expérimentateur invite l'auditeur à se familiariser avec l'outil de segmentation. Il s'agit de signaler « à la volée » la limite des blocs de parole en appuyant sur la barre espace du clavier de l'ordinateur. L'expérimentateur insiste sur le fait que la réaction doit être très rapide

⁵ Perruchet, Tyler, Galland et Peereman (2004) ont souligné la présence d'un biais dans la parole obtenue par la synthèse avec MBROLA. Les occlusives non voisées y sont précédées d'un espace silencieux plus long que celui observé en langage naturel. La durée du silence est proportionnelle à la longueur du phonème considéré.

⁶ Les programmes sont rédigés en langage tcl et exécutables avec l'interpréteur tcl/tk sous système X. Les divers fichiers de données (réponses individuelles et résumés) sont édités et gérés par le logiciel Praat.

⁷ « bonjour, monsieur » ; « bonjour monsieur » ; « le très grand cheval de bois » ; « le très grand, cheval de bois ».

pour ne pas empiéter sur l'écoute de ce qui suit. Vient ensuite une phase de mesure du temps de réaction moyen de l'auditeur sur la base de ses réponses à l'audition d'une série de 10 bips délivrés à intervalles irréguliers. Puis l'auditeur est invité à segmenter les deux brèves séquences énoncées par le locuteur de l'échantillon expérimental. Il segmente ensuite l'une des quatre versions de l'échantillon expérimental. L'expérience dure 15 minutes environ.

Le programme qui gère l'expérience enregistre les temps de réponse. Pour chaque réponse donnée, il mesure la durée écoulée entre le début du signal et l'appui sur la touche, puis calcule une valeur corrigée en fonction du temps de réaction moyen. Les deux valeurs (brute et corrigée) sont enregistrées dans un fichier individuel de réponse.

Plan de recueil

Chaque participant a segmenté une seule des versions de l'échantillon expérimental.

Participants

L'analyse perceptive des quatre versions a nécessité la participation de 60 auditeurs (15 participants x 4 versions). Les participants sont des adultes francophones de l'Université de Provence, sans connaissances expertes en prosodie.

Prédictions

Les auditeurs ont pour consigne de segmenter les paroles entendues en séquences (ou blocs) de mots semblant former un tout. En premier lieu, du fait qu'il traitent en un seul passage le signal entendu, on s'attend à ce que la segmentation des quatre versions soit toujours plus grossière que celle effectuée par un expert lors de l'écoute réitérée de la version originale. Il est attendu des participants qu'ils signalent seulement les démarcations prosodiques perceptivement les plus fortes (hypothèse 1). En second lieu, bien que la consigne invite les auditeurs à se focaliser sur "la musique de la parole", on s'attend à ce que les groupements qu'ils effectuent dans le flux sonore varient en fonction du type d'information disponible pour le traitement psycholinguistique de chaque version (hypothèse 2). La version originale et la version re-synthétisée, outre l'information prosodique offrent un signal de parole qui autorise l'identification des items lexicaux et grammaticaux, donnant ainsi accès au contenu propositionnel des énoncés. La version filtrée et la version saltra~dzajn n'autorisent pas l'accès au contenu propositionnel des énoncés. Cependant la version saltra~dzajn offre un signal de parole fréquemment complet qui autorise l'individualisation de formes lexicales non interprétables, ce que ne permet pas la version filtrée. La version filtrée devrait être, des quatre versions, celle où les auditeurs ont le plus de difficulté à signaler sur la base de l'information disponible, des mots ou des groupes de mots qui vont ensemble. En revanche la version saltra~dzajn qui libère l'auditeur de toute information lexico-syntaxico-sémantique devrait être celle qui permet le mieux l'individualisation de blocs de parole sur la base de l'information prosodique qu'elle dispense. L'ensemble devrait se traduire au niveau de la performance de chaque groupe par un nombre décroissant de segments individualisés depuis la version saltra~dzajn jusqu'à la version filtrée, en passant par les versions originale et re-synthétisée.

En outre, l'analyse comparée des groupements effectués dans chaque version devrait permettre de dégager des points de convergence entre les segmentations qui peuvent être considérés comme principalement fondés sur les caractéristiques prosodiques du signal. Le résultat des analyses acoustiques en ces points devrait permettre de préciser quelles sont les variations des paramètres analysés prises en compte lors du signalement des blocs.

RESULTATS

N^ombre de réponses par participant

Sur l'ensemble de l'expérience, les participants ont donné en moyenne un peu moins de 20 réponses lors de l'audition des différentes versions de l'échantillon expérimental (moyenne = 18,9 ; écart-type = 1,7) (Tableau 2).

#Tableau 2 : nombre de réponses par participant selon les versions

versions	Originale	Filtrée	Re-synthétisée	Saltra~dzajn
moyenne	20,27	16,5	19	20
Ecart type	5,9	4,3	5,1	4,6

Une analyse de la variance destinée à tester l'effet des versions avec le *nombre de réponses* données par chaque participant pour variable dépendante ne montre pas d'effet significatif. Les données recueillies ne permettent pas de conclure à une différence quant à la quantité de réponses fournies par chacun selon les versions.

Localisation des réponses

Les valeurs corrigées des temps de réponses (TR) ont été utilisées pour situer en quels points du signal (ou sites) chaque réponse a été donnée. Pour chaque version, les TR de chacun des 15 participants (recueillis dans des fichiers individuels) ont été reportés dans un fichier unique rassemblant les résultats du groupe.

Les grilles de dépouillement des réponses données pour chaque version ont été conçues comme suit. Au moyen du logiciel Praat, le signal de chaque version a été découpé a priori en intervalles de durée variable. Ces intervalles ont pour limite la fin des mots supports des contours intonatifs identifiés par Portes (2004) dans la version originale ou la fin d'un long espace silencieux, soit au total 46 intervalles. Pour chaque version, Praat a permis de préciser la valeur des bornes des intervalles définis dans la grille, mesurée depuis l'origine du signal.

La durée du signal jusqu'à la limite droite des intervalles établis pour chaque version a été utilisée comme référence pour situer les TR individuels. Lorsque, pour un participant donné, plusieurs mesures concernent le même intervalle, ces mesures sont ordonnées dans les limites de l'intervalle par ordre de valeur croissante. Au terme du dépouillement, pour chaque version, les réponses des 15 participants ainsi rassemblées et alignées selon leur valeur entière (exprimée en secondes), font apparaître des concordances et des discordances dans les réponses.

Base de données

Afin de résumer pour chaque version l'information concernant la localisation sur le signal des réponses données, des moyennes inter sujets ont été calculées lorsque les réponses étaient concordantes. Le critère de concordance était une variation admise inférieure à la seconde entre deux TR. Les sites résumant la performance du groupe reflètent le plus souvent la réponse de plusieurs auditeurs, mais ils reflètent parfois la réponse d'un seul auditeur. Le Tableau 3 présente les caractéristiques des réponses pour chaque version.

#Tableau 3 : caractéristiques des réponses selon les versions

versions	Originale	Filtrée	Re-synthétisée	Saltra~dzajn
Nombre de sites	66	54	68	61

Nb maximum de participants ayant contribué à l'établissement d'un site	13	12	11	11
Nombre de sites indiqués par un seul participant	15	11	10	16

Globalement, le *n`mbre sites* est égal en moyenne à 62,25 (écart-type = 6,2). L'examen de la *pr`p`rti`n d'auditeurs* ayant contribué à l'individualisation de chaque site montre que, quelle que soit la version, un site donné n'a jamais été établi sur la base des réponses des 15 auditeurs du groupe (réponse unanime). Par ailleurs, les sites indiqués par un seul participant (réponses isolées) représentent entre 15 et 26 % des sites selon les versions. Une analyse de la variance destinée à tester l'effet des versions avec le *n`mbre de rép`nses is`lées* pour variable dépendante ne montre cependant pas d'effet significatif et ne permet pas de conclure à une différence concernant la quantité de réponses isolées selon les versions.

Pour finir la description, l'examen du *n`mbre de sites* dégagés pour chaque version (qu'il soit tenu compte ou non des sites indiqués par un seul participant), montre que ce nombre varie d'une version à l'autre : qu'il se révèle minimal pour la version filtrée et maximal pour la version re-synthétisée. Pour chaque version, ces sites constituent la base de données des analyses qui suivent.

Sites maj`ritaires

Pour les quatre versions, la base de données met sur le même plan aussi bien les sites établis grâce aux réponses de plusieurs participants, que les sites signalés par un seul auditeur du groupe. Une première analyse a consisté à ne considérer que les sites indiqués au moins par 8 auditeurs sur 15 (réponse majoritaire). Pour apprécier l'accord entre les auditeurs, un indice a été calculé qui rapporte les accords majoritaires observés, aux accords majoritaires possibles étant donné les 46 sites possibles de réponse définis dans la grille de dépouillement. En suivant la procédure décrite dans Passoneau & Litman (1993), un pourcentage global d'accord (qui prend en compte aussi bien la présence que l'absence majoritaire de réponses pour un site possible) et un pourcentage d'accord concernant la présence de réponses majoritaires pour les sites indiqués ont été calculés. Le tableau 4 présente le nombre de sites majoritaires et les indices d'accord établis pour chaque version.

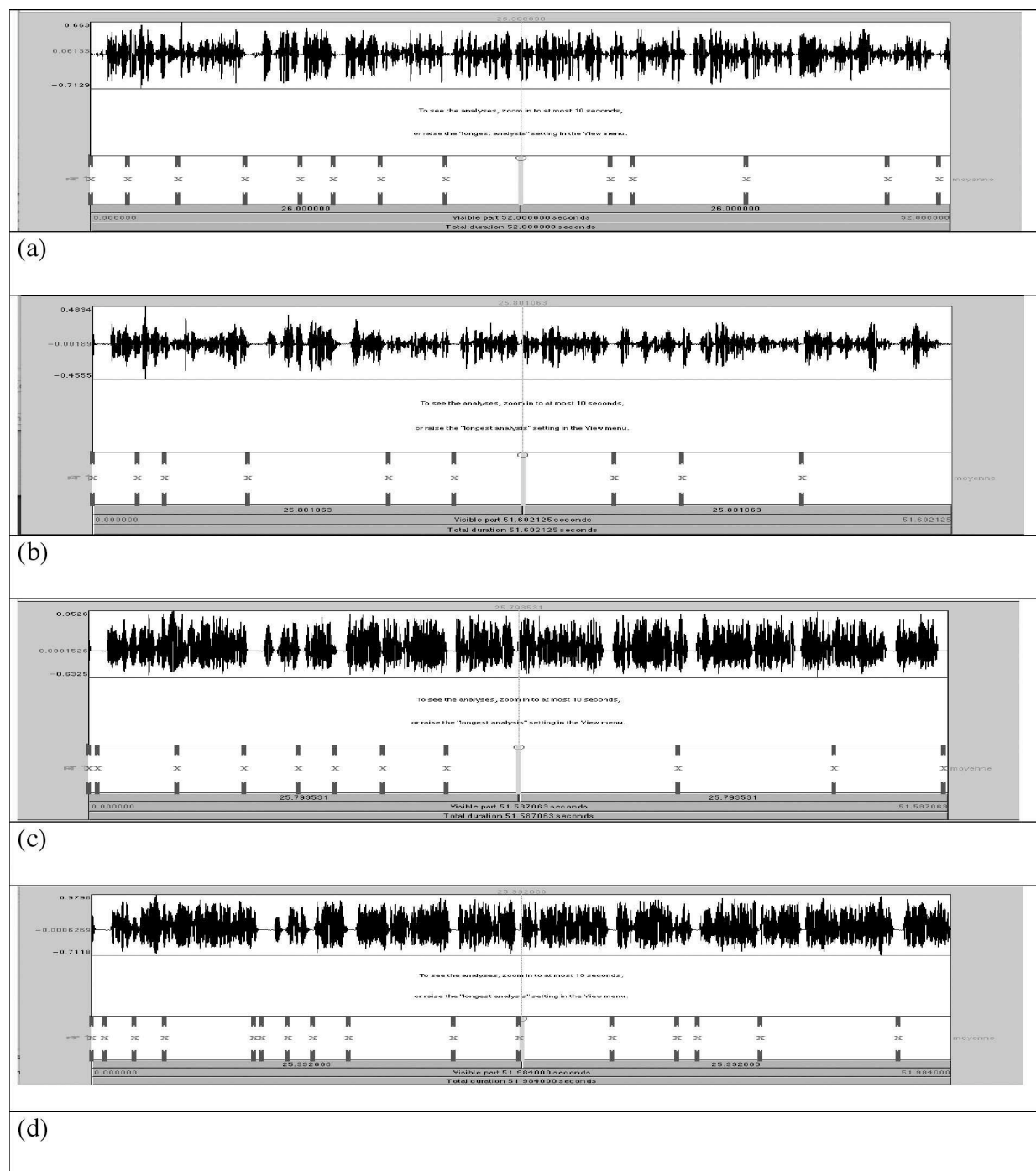
#Tableau 4 : sites majoritaires : accord entre les auditeurs selon les versions

versi`ns	Originale	Filtrée	Re-synthétisée	Saltra~dzajn
Nb de sites maj`ritaires	12	8	10	15
% gl`bal d'acc`rd	73,62	78,99	74,93	79,5
% d'acc`rd p`ur les sites indiqués	67,9	69,74	70	77,3

Globalement, le nombre de sites majoritaires est en moyenne de 11,25 sites pour 46 sites possibles de réponse (24,5 % ; écart-type = 3) soit environ 18% des sites de la base. Examiné en fonction de chaque version, ce nombre est maximal pour la version saltra~dzajn (15 sites). Il est minimal pour la version filtrée (8 sites). Une analyse de la variance conduite avec le *n`mbre de sites maj`ritaires* pour variable dépendante montre un effet significatif des versions ($F(3,56) = 8,331$; $p < .0001$). Une comparaison par paires, effectuée au moyen du test Newman-Keuls, montre (au seuil .01) que le nombre de sites majoritaires de la version saltra~dzajn est significativement supérieur à celui des versions re-synthétisée et filtrée.

Pour toutes les versions, l'indice d'accord global (76,76 % en moyenne ; écart-type = 3) est toujours plus élevé que l'indice d'accord pour les seuls sites indiqués (71,24 % en moyenne ; écart-type = 4,2). En outre pour les deux indices, *l'accord sur la segmentati`n* est maximal pour la version saltra~dzajn et minimal pour la version originale.

A ce niveau de sélection des données (réponses majoritaires), l'examen de la localisation des sites fait apparaître des segmentations du signal différentes pour chaque version. La figure 1 présente la *répartiti`n de ces sites sur le signal* correspondant à chaque version.



#Figure 1 : répartition des sites majoritaires pour chaque version.

C`mparais`n des segmentati`ns

L'analyse comparée de la segmentation des différentes versions de l'échantillon expérimental a été effectuée en prenant pour référence le résultat de deux analyses prosodiques de la version originale par des experts. Ceux-ci ont pu écouter et réécouter le matériel autant que nécessaire pour parachever leur analyse. Le premier résultat est celui de l'analyse auditive des contours intonatifs réalisée par Portes (2004). Le second provient de la tâche effectuée par trois autres experts qui, indépendamment, ont évalué la force de chaque frontière de mots au moyen d'une

échelle en 4 points, où “1” signifie “frontière très faible” et “4” “frontière très forte”. Cette pondération s’inspire de la notation des niveaux de jointure en pratique dans le système de transcription ToBI (“Tone and Break Indices” ; Silverman, Beckman, Pitrelli, Ostendorf, Wigham, Price, Pierrehumbert & Hirschberg, 1992) qui permet d’évaluer le groupement et la démarcation prosodiques des mots dans une séquence de parole. A l’issue de cette seconde analyse, 17 frontières de mots ont reçu à l’unanimité le poids 3 ou 4. Dans la transcription intégrale en annexe (ligne b), les mots concernés à leur frontière droite par ces poids consensuels sont indiqués en italiques gras.

Analyse de la version originale par les experts

L’analyse des contours intonatifs exclut par principe les hésitations, les phatiques et les pauses qui sont conçus comme des traces de l’activité d’encodage. Alors que la pondération des jointures de mots de cette même version concerne tout le matériau segmental. Il ressort de cette seconde analyse que les disjonctions jugées unanimement fortes ou très fortes par les trois experts concernent aussi bien des mots supports d’un contour intonatif que des phatiques (« heu »).

La Figure 2 présente la distribution des contours selon les catégories intonatives (1^{ère} analyse) et la répartition des frontières droites de mots pondérées 3 ou 4 par les trois experts (2^{ème} analyse) selon ces mêmes catégories.

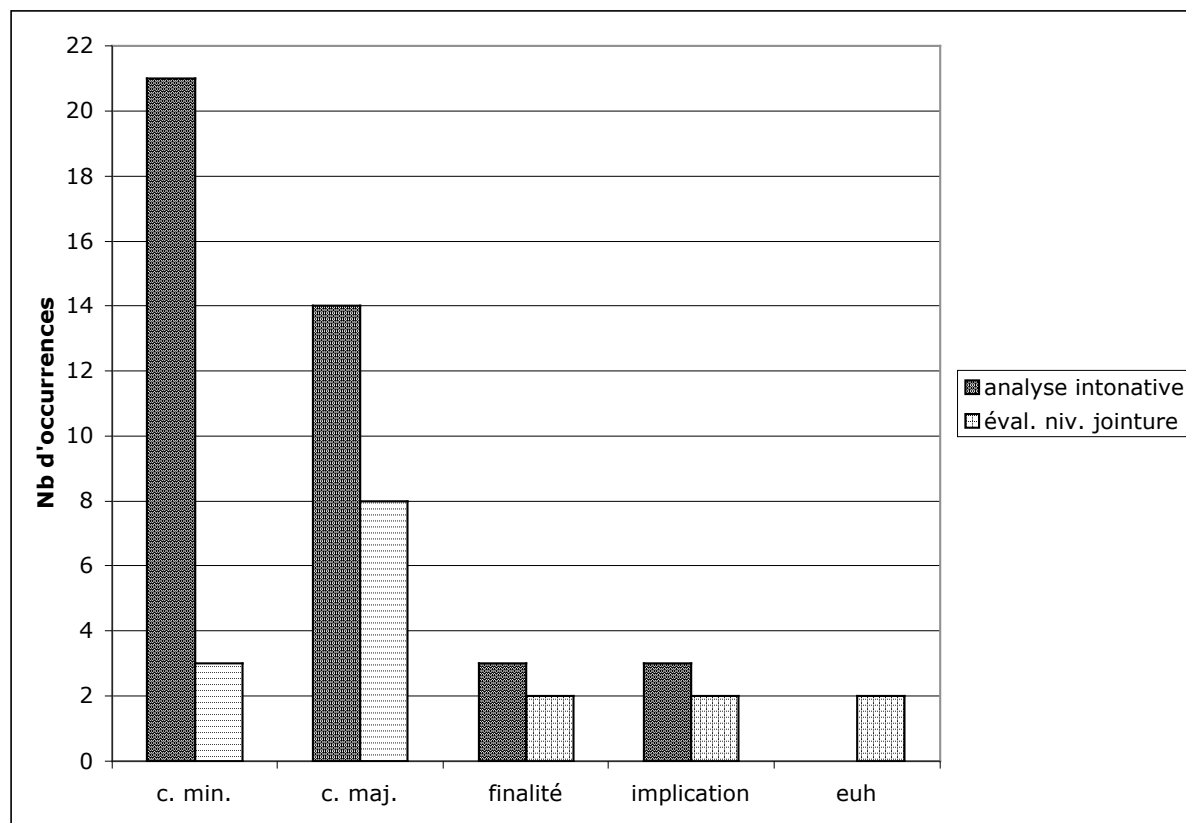


Figure 2 : version originale : distribution des contours et des frontières de mots jugées fortes ou très fortes en fonction des catégories intonatives.

La superposition des résultats des deux analyses montre que, deux fois sur trois, les experts ont attribué un poids 3 ou 4 à des frontières de mots supports de contours de finalité⁸ ou d'implication⁹. Ils ont également attribué ces poids à plus de la moitié des frontières de mots supports d'un contour de continuité majeure¹⁰ (57,1%). En revanche ils n'ont attribué ces poids qu'à 14% des frontières de mots supports d'un contour de continuité mineure¹¹.

Ainsi, la pondération des frontières de mots par les trois experts, qui rend compte non seulement de certaines frontières intonatives, mais également d'éléments phatiques, peut-elle servir de base de référence pour vérifier l'hypothèse selon laquelle lors du traitement du signal en temps réel, les auditeurs peuvent indiquer sur le champ les démarcations prosodiques perceptivement jugées les plus fortes (hypothèse 1).

Comparaison de la segmentation des quatre versions par les tout-venant

La comparaison du résultat de la segmentation de chaque version par les auditeurs tout-venant a été effectuée en changeant le grain d'examen des données de la base. Les sites sélectionnés pour cette analyse sont ceux issus des *rép`nses d'au m`ins 6 auditeurs sur 15* (soit 40 % des auditeurs¹²). Contrairement à la sélection des données issue de l'application du critère majoritaire, les sites sélectionnés selon ce nouveau critère représentent pour chaque version un effectif proche du nombre de frontières de mots de la version originale pondérées 3 ou 4 par les experts. Cet effectif est en moyenne égal à 19 (écart type = 3,9). Les sites sélectionnés représentent globalement 30,5% des sites qui figurent dans les données de la base et ils représentent 41,3% des 46 sites de la grille. Le Tableau 5 présente pour chaque version, le nombre de sites ainsi sélectionnés et les indices d'accord correspondants (la segmentation de chaque version établie selon ce critère figure en détail en annexe, lignes c à f).

#Tableau 5 : sites établis sur la base des réponses d'au moins 6 auditeurs sur 15

versions	Originale	Filtrée	Re-synthétisée	Saltra~dzajn
Nombre de sites	20	17	15	24
% d'accord pour les sites indiqués	60	60,33	57,65	65,22
Nombre de sites focalisés à l'intérieur d'un mot	4 20%	5 29,4%	8 53,4%	3 12,5%

Examiné en fonction des versions, le nombre de sites est maximal pour la version saltra~dzajn (24 sites) et minimal pour la version re-synthétisée (15 sites). Une analyse de la variance conduite avec le *nombre de sites* pour variable dépendante montre un effet significatif des versions ($F(3,56) = 7,600$; $p < .0002$). Une comparaison par paires, effectuée au moyen du

⁸ Un contour de **finalité** a un mouvement descendant d'un niveau moyen ou haut abaissé dans le registre du locuteur jusqu'au niveau bas et signale la fin de l'énoncé.

⁹ Le contour d' **implicatif** correspond à une fin de proposition. Il est ascendant de moyen à haut sur la dernière syllabe qui précède la frontière et redescend immédiatement sur la même syllabe à une valeur beaucoup plus basse.

¹⁰ La courbe de la **continuité majeure** est montante. Elle se distingue du contour montant d'implication par l'alignement du maxima de F0 dans la deuxième moitié de la voyelle (Portes, 2004).

¹¹ La courbe de la **continuité mineure** se situe dans les niveaux de hauteur moyen ou haut abaissé. Elle n'a pas de direction fixe. Elle peut aussi bien descendre que monter. L'écart entre le minimum et le maximum de la courbe intonative est moindre que pour la continuité majeure.

¹² Ce critère est comparable à celui employé par Lehiste (1979).

test Newman-Keuls, montre (au seuil .01) que le nombre de sites de la version saltra~dzajn est significativement plus élevé que celui des versions re-synthétisée et filtrée.

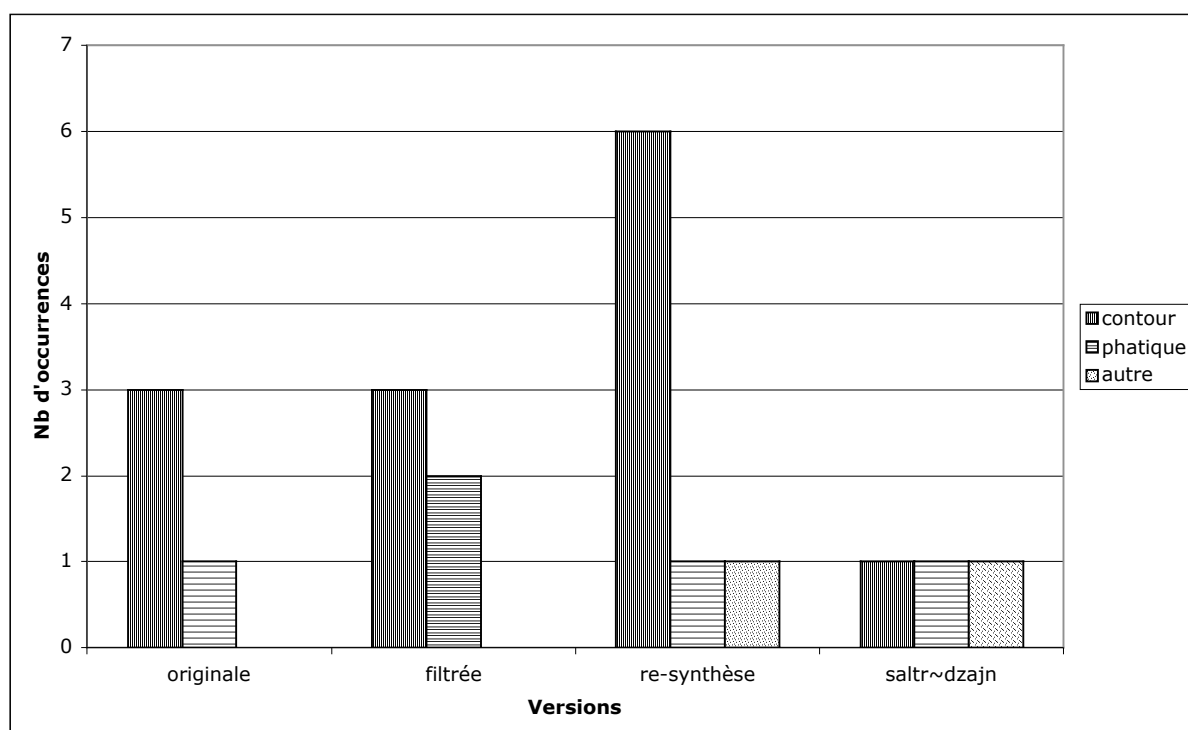
Pour les quatre versions, l'accord global moyen entre les auditeurs pour les sites indiqués est similaire et égal en moyenne à 60,8 %. Examiné en fonction des versions, *l'accord moyen entre les auditeurs* est maximal pour la version saltra~dzajn (65,22 %) et minimal pour la version re-synthétisée (57,65 %).

En résumé, l'examen des caractéristiques des segmentations ci-dessus (Tableau 5), comparé à celui des segmentations établies selon le critère majoritaire (Tableau 4), souligne comme pour ces dernières un plus fort degré d'accord pour les sites de la version saltra~dzajn et un découpage plus fourni de cette version que celui des versions re-synthétisée et filtrée. Ce qui va dans le sens de l'hypothèse 2.

Précision de la localisation

La localisation d'un site résulte du calcul d'une durée moyenne fondée sur six réponses individuelles. Reportée sur le signal, cette durée peut se situer à la jointure de deux mots contigus ou bien dans un mot. La *précision de la localisation* des sites a été évaluée en comptant le nombre de fois où la durée reportée sur le signal se situe dans un mot (site inclus). Globalement (Tableau 5), la segmentation de la version saltra~dzajn est plus précise que celle de la version re-synthétisée : la version saltra~dzajn comporte 3 sites inclus (ce qui représente 12,5% des sites) contre 8 pour la version re-synthétisée (soit 53,4% des sites).

Le mot récepteur du site inclus peut être un item lexical support d'un contour intonatif ou bien un item lexical quelconque ou encore un élément phatique. La Figure 3 présente pour chaque version la distribution des sites inclus en fonction de ces trois catégories.

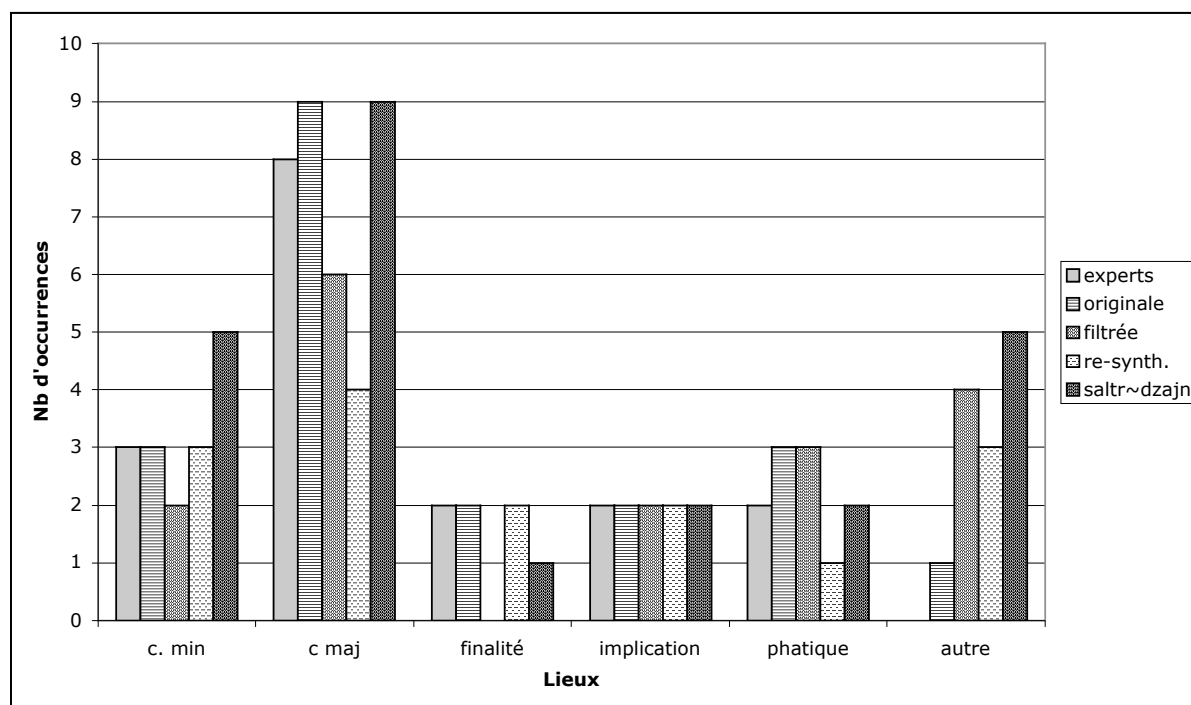


#Figure 3 : distribution des sites inclus dans chaque version en fonction de leur localisation
[Légende : Contour : mot support d'un contour intonatif ; Autre : mot ne supportant pas un contour intonatif ; Phatique : élément phatique].

A l'exclusion de la version saltra~dzajn pour laquelle la distribution est uniforme, l'examen de la distribution des sites inclus montre que sont concernés, à une exception près, des mots supports d'un contour intonatif ou des éléments phatiques. L'inclusion dans un mot support de contour intonatif est très fréquente en version re-synthétisée. Elle a lieu entre la voyelle finale accentuée et la coda consonantique. Compte tenu du fait qu'en français, la voyelle finale accentuée est porteuse de l'essentiel de l'information concernant la frontière intonative (allongement et type de contour), l'inclusion d'un site dans un mot support de contour intonatif est interprétable comme un cas de réponse anticipant la frontière.

L'analyse des réponses

Compte non tenu des anticipations, les sites établis pour chaque version peuvent correspondre à la limite droite d'un élément phatique ou d'un contour intonatif ou à d'autres composants du signal. La figure 4 présente pour les quatre versions la distribution des sites en fonction de ces lieux. En ce qui concerne les contours intonatifs, la catégorie est subdivisée en tenant compte du type de contour. Dans la figure 4, la distribution des frontières de mots pondérées (3 ou 4) par les experts selon ces mêmes lieux est donnée pour référence.



#Figure 4 : répartition des frontières de mots jugées fortes et très fortes (version originale) et des sites établis pour chaque version (accord : 6 / 15), en fonction de leur localisation : à la limite droite d'un contour intonatif (catégories : « continuité mineure », « continuité majeure », « finalité, implication »), d'un élément « phatique », ou d'autres composants du signal (« autre »).

Dans l'ensemble ainsi qu'attendu, les distributions issues du découpage des quatre versions de l'échantillon par les auditeurs tout-venant, sont très proches de la distribution des frontières de mots jugées fortes et très fortes par les experts. Les sites correspondent principalement à certaines frontières intonatives ou à des éléments phatiques. Dans chacun de ces lieux le nombre de sites est proche du nombre de frontières de mots jugées fortes ou très fortes par les experts.

La segmentation des quatre versions est en accord sur le nombre de sites relevant de la frontière de contours d'implication ; celle des versions originale et re-synthétisée est en accord sur le nombre de sites relevant de frontières de finalité et de continuité mineure ; et celle des versions originale et saltra~dzajn est en accord sur le nombre de sites relevant de frontières de continuité majeure.

Il existe aussi un certain nombre de cas observés hors des lieux signalés par les experts. Ils concernent surtout les versions dérivées de la version originale : réaction à un bruit sur le signal, signalement retardé d'une frontière ou de la fin ou du début d'un mot qui n'est pas un support de contour intonatif (7,9 % des sites).

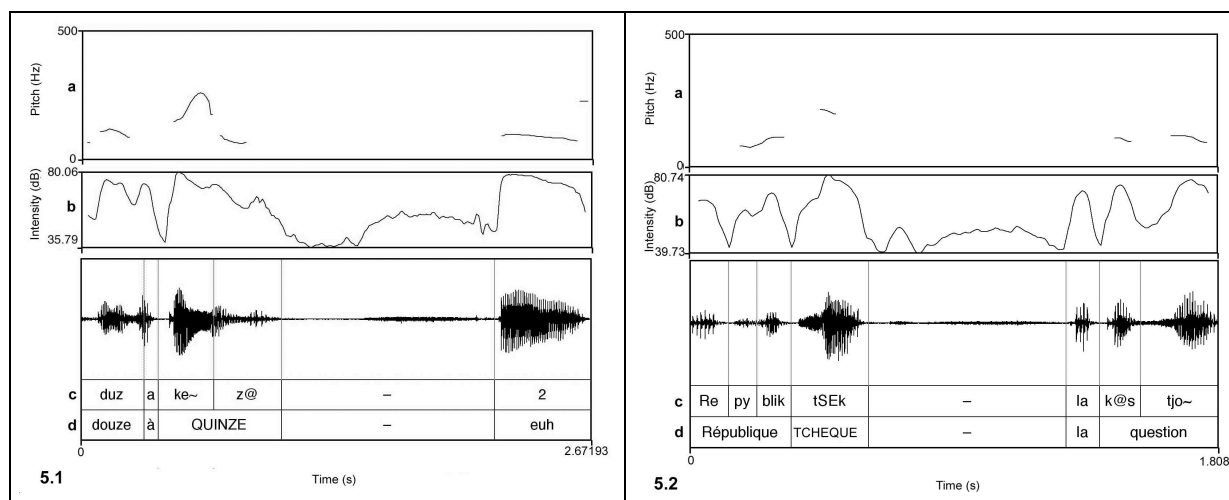
En résumé, lors de la segmentation des versions proposées, les auditeurs tout-venant ont principalement indiqué des points du signal correspondant, selon l'évaluation experte de la version originale, à des démarcations prosodiques entre mots jugées fortes ou très fortes et qui le plus souvent correspondent à des frontières intonatives. Ce qui va dans le sens de l'hypothèse 1.

Caractéristiques prosodiques des sites communs aux quatre versions

La récapitulation des analyses auditives de l'échantillon original et de ses différentes versions qui figure en annexe, met en regard le résultat des segmentations réalisées par les auditeurs tout-venant (sites issus des réponses d'au moins 6 auditeurs sur 15). Cette disposition permet de repérer les zones du signal où les réponses des quatre versions convergent. La segmentation des quatre versions présente huit zones de convergence. Ce sont dans ces zones que les auditeurs sont le plus susceptibles d'avoir exploité principalement la prosodie pour répondre. L'analyse qui suit tente de préciser, par l'examen de ces zones, les caractéristiques prosodiques sur lesquelles les auditeurs se sont effectivement fondés.

De façon générale, les phénomènes prosodiques impliqués dans la structuration de la parole spontanée et du discours peuvent survenir au niveau même d'une frontière (patrons prosodiques, pauses) ou concerner l'ensemble du domaine délimité par deux frontières (variations du débit, du registre ou de la pente intonative). L'examen des zones de convergence se limite ici aux seuls phénomènes acoustico-prosodiques qui surviennent dans l'environnement immédiat des réponses. Leur description se fonde sur l'analyse automatisée du signal de la version originale selon divers paramètres prosodiques dont la variation de la fréquence fondamentale (F0), de l'intensité et de la durée (pausale, segmentale et syllabique).

Les réponses convergent dans deux zones où l'analyse intonative signale la présence de contours d'**implication** (figures 5). Il s'agit des deux zones « ... à QUINZE » et « ... la république TCHEQUE » où l'analyse des sites majoritaires montrait des points d'accord dans la segmentation des versions.

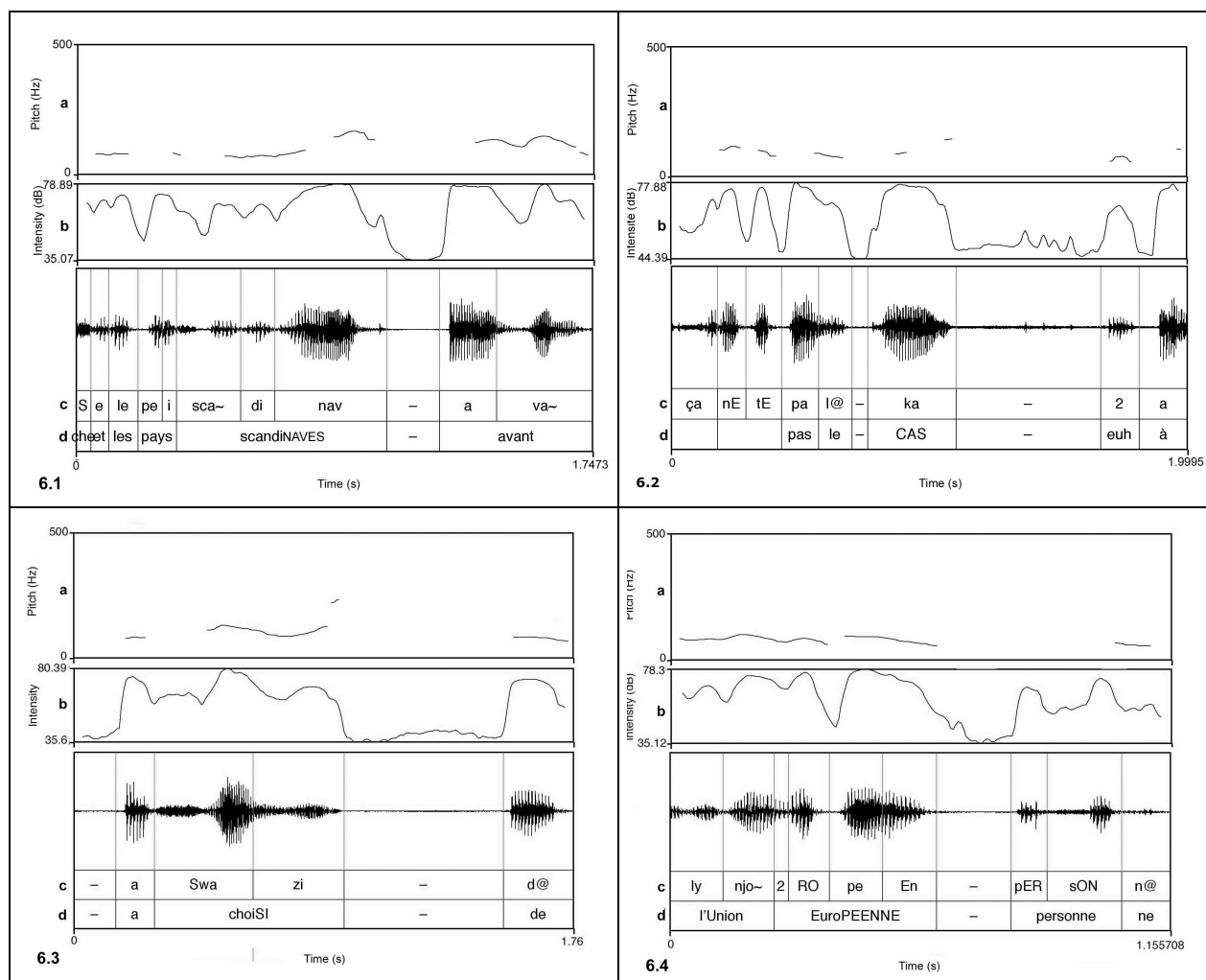


#Figures 5 : contours d'implication [Légende :(a) F0 (Hz) ;(b) intensité globale (dB) ;(c) transcription SAMPA ; (d) transcription orthographique].

Ainsi que Portes (2004) a pu le préciser par l'analyse du corpus dont est extrait l'échantillon expérimental, ces contours présentent une élévation très rapide de la F0 sur la syllabe qui précède la frontière intonative. La valeur de la F0 y atteint la partie haute du registre du locuteur dans la première moitié de la voyelle. Elle redescend immédiatement pour atteindre la partie basse du registre. Ici les deux contours sont suivis d'un silence.

La réponse donnée pour [k e~z] (figure 5.1) qui est suivi d'un long silence (1104 ms), survient avant la fin de la syllabe pour les versions filtrée et re-synthétisée et dès la fin de l'émission de la syllabe dans la version originale. Dans la version saltra~dzajn, une réponse a lieu avant la fin de la syllabe et l'autre immédiatement après. La réponse donnée pour [tSEk] (figure 5.2) qui est suivi d'un silence (673 ms), survient avant la fin de la syllabe dans la version re-synthétisée et dès la fin de la syllabe dans les trois autres versions.

Les réponses convergent dans quatre autres zones où l'analyse intonative signale la présence de contours de **continuité majeure** (figures 6). Leur description prototypique fait état d'une élévation très rapide de la valeur de la F0 sur la syllabe précédant la frontière, qui atteint son maximum dans la deuxième moitié de la voyelle. Trois cas se conforment à cette description.



#Figures 6 : contours de continuité majeure [Légende : (a) F0 (Hz) ; (b) intensité globale (dB) ; (c) transcription SAMPA ; (d) transcription orthographique].

Pour la syllabe [nav] dans la zone «... les pays scandiNAVES » (figure 6.1), la F0 atteint son maximum dans la partie haute du registre du locuteur. Un silence de 117 ms suit la fin du contour. Dans cette zone, la réponse survient avant la fin de la syllabe dans les versions originale, filtrée et re-synthétisée et vers le milieu du silence qui la suit dans la version saltra~dzajn.

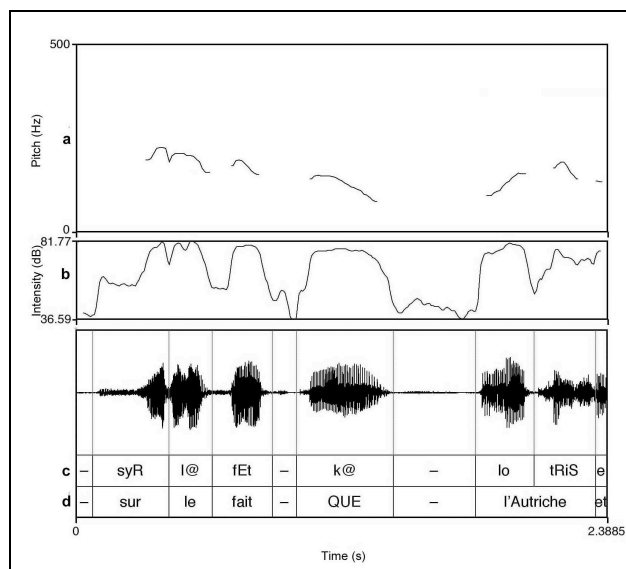
La réponse dans la zone « ... pas le CAS » (figure 6.2) survient dans les quatre versions lors du silence (556 ms) qui suit la syllabe [ka] (lors du premier tiers dans la version saltra~dzajn, du tiers dans les versions filtrée et re-synthétisée et à la toute fin dans la version originale).

La réponse donnée dans la zone « ... a choiSI » (figure 6.3), où [zi] est suivi de 561 ms de silence, survient vers le milieu de ce silence dans les versions originale et saltra~dzajn. Mais elle survient dans les deux autres versions à la limite entre [a] émis avec une forte intensité et l'initiale de la syllabe qui précède [zi].

La quatrième zone « ... l'union européENNE » (figure 6.4) se démarque de la description prototypique. la syllabe [En] présente dans la partie basse du registre du locuteur un contour descendant suivi de 233 ms de silence. La réponse survient immédiatement après la fin de la

syllabe [En] dans les versions originale, re-synthétisée et saltra~dzajn et à la fin du silence qui suit dans la version filtrée.

Les réponses convergent dans une zone où l'analyse intonative signale la présence d'un contour de **continuité mineure** descendant (figure 7).

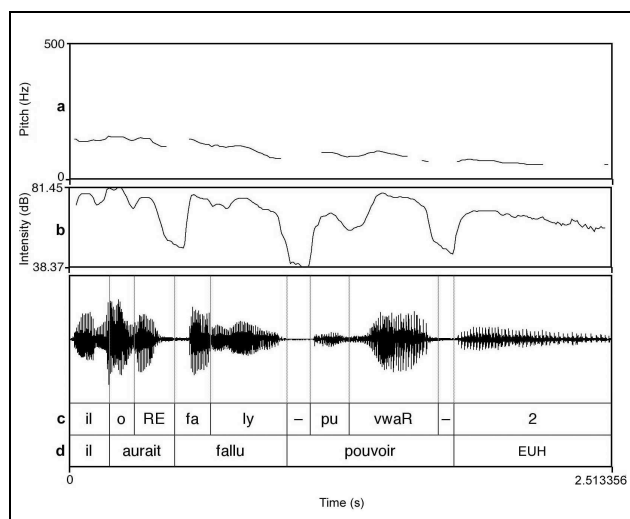


#Figure 7 : continuité mineure à contour descendant [Légende : (a) F0 (Hz) ; (b) intensité globale (dB) ; (c) transcription SAMPA ; (d) transcription orthographique].

Dans la zone « sur le fait QUE », la valeur de la F0 sur la dernière syllabe chute très rapidement de la partie moyenne à la partie la plus basse du registre du locuteur. Suit un espace silencieux de 369 ms.

La réponse survient avant la fin de la syllabe [K@] dans la version re-synthétisée, dès sa fin dans la version originale, au milieu du silence qui suit dans la version saltra~dzajn et à la fin de ce silence dans la version filtrée.

Enfin les réponses convergent dans une zone qui correspond à la présence d'un élément phatique « ...pouvoir HEU » (figure 8)



#Figure 8 : élément phatique [Légende : (a) F0 (Hz) ; (b) intensité globale (dB) ; (c) transcription SAMPA ; (d) transcription orthographique.

Pendant la demi seconde que dure la voyelle [2], le contour est plat : la valeur de la F0 se maintient dans la partie la plus basse du registre du locuteur et s'accompagne d'une évolution de l'intensité progressivement décroissante. La réponse donnée pour [2] survient, dans toutes les versions, avant la fin de l'élément phatique.

En résumé, l'examen des caractéristiques prosodiques de l'environnement des sites communs aux versions fait ressortir que la présence d'un silence consécutif à l'émission d'une séquence de parole continue ne contribue pas toujours à la décision des auditeurs. Quelle que soit la version, le silence n'a pas contribué à la décision dans deux cas. Lors de l'audition des contours d'implication, les réponses sont survenues en cours ou immédiatement à la fin de l'audition de la syllabe porteuse du mouvement circonflexe de la F0 qui implique la partie haute du registre du locuteur. Lors de l'audition de l'élément phatique les réponses sont survenues en cours d'audition de la voyelle dont le contour est plat dans la partie la plus basse du registre. En revanche, la présence d'un silence subséquent peut avoir participé à la décision dans le cas des contours de continuité. Sur la base des occurrences examinées ici, cette participation semble potentiellement différente selon les versions sans que pour autant des régularités soient apparues.

CONCLUSION

De même que les tâches comportementales et les procédures d'annotation de corpus, l'outil de segmentation testé ici exige un traitement de la prosodie en attention focalisée. Dans la communication parlée, le traitement de la prosodie est effectué parallèlement à la mise en oeuvre de processus de traitement spécifiques, comme les processus syntaxiques et lexicaux, mais aussi de processus cognitifs généraux (inférentiels et mnésiques) responsables de l'interprétation. Il semble vraisemblable de considérer que dans l'échange de paroles, les ressources attentionnelles allouées aux processus de décodage grammatical et lexical ainsi qu'à l'interprétation du discours sont plus importantes que les ressources allouées au traitement de la prosodie. Aussi, du point de vue de sa validité écologique, par les jugements prosodiques qu'il

sollicite, l'emploi de l'outil de segmentation s'inscrit-il clairement dans un cadre métalinguistique plutôt que communicatif. Mais à la différence des autres tâches, son emploi peut permettre d'approcher le processus d'intégration perceptive de l'information prosodique en ligne, tel qu'il se produit d'ordinaire.

Sachant que la compréhension se fonde à la fois sur des indices relevant de l'organisation prosodique, syntaxico-sémantique et textuelle du discours, le maniement de l'outil de segmentation a été examiné lors du traitement de trois versions dérivées d'un échantillon de paroles produites dans un cadre communicatif authentique. La version originale et la version re-synthétisée de cet échantillon offrent la totalité de l'information qui permet l'accès au contenu des énoncés et à leur interprétation. La version filtrée ainsi que la version re-synthétisée avec conversion de phonèmes (version saltra~dzajn), en dégradant les indices permettant l'accès au contenu, dispensent les auditeurs du traitement lexico-syntaxico-sémantique. Elles évitent donc de proposer un matériel d'ordinaire composite qui peut faire courir le risque aux participants « d'entendre » une frontière prosodique à la fin de chaque proposition. Cependant la version filtrée rend difficile le repérage de formes pouvant constituer des mots, ce qui n'est pas le cas de la version saltra~dzajn. Eu égard à la consigne donnée ici aux participants, cette dernière version devait être celle qui leur permet le mieux de se focaliser sur l'individualisation de séquences de parole sur la base de l'information acoustico-prosodique disponible et donc celle qui distord le moins les jugements prosodiques.

Les données collectées auprès de participants tout-venant, ont été analysées et mises en regard du résultat de deux analyses expertes de l'échantillon original : celle de ses contours intonatifs et celle des niveaux de jointure entre les mots.

La comparaison de la segmentation de l'échantillon original par les tout-venant aux résultats des analyses expertes va dans le sens attendu. En faisant varier le grain de l'analyse selon le nombre de participants en accord sur l'indication d'un point donné du signal, il s'est avéré possible de localiser précisément les variations acoustico-prosodiques jugées perceptivement les plus fortes pour les auditeurs (les points signalés par au moins 40 % des auditeurs sont interprétables dans plus de 92 % des cas).

Ainsi qu'attendu également, des trois versions dérivées de l'échantillon original, la version re-synthétisée avec conversion de phonèmes est non seulement celle qui permet, à égalité avec la version originale le repérage des contours de continuité majeure, mais aussi celle qui permet de localiser le plus grand nombre de contours de continuité mineure.

Enfin l'examen des points de convergence (points signalés à l'audition de toutes les versions par au moins 40 % des auditeurs) a permis d'approcher le jeu tantôt exclusif et tantôt conjugué de diverses variations prosodiques pour leur signalement. C'est ainsi qu'il est apparu que les mouvements circonflexes amples et rapides de la F0 (contour d'implication) ou au contraire son état stationnaire dans la partie basse du registre (phatique) sont signalés indépendamment du silence qui se trouve à leur suite, alors que ce n'est pas le cas pour les contours de continuité. Les caractéristiques prosodiques des contours d'implication et des éléments phatiques observés ici forment une partie de l'ensemble réduit de patrons intonatifs décrits dans les paroles que la mère adresse à son bébé nouveau né dès les premiers jours (Fernald & Simon, 1984) et auxquels ce dernier réagit également très tôt. Ces mêmes caractéristiques prosodiques se retrouvent ultérieurement dans le langage maternel adressé au bébé pour attirer son attention et lui parler des objets et des événements du monde environnant (Colas, 1999). Ces observations vont dans le sens de l'idée que ces caractéristiques fonctionnent comme des marqueurs attentionnels. Et

sont un moyen de mobiliser une attention plus soutenue chez l'auditeur afin de faciliter ou d'améliorer le fonctionnement d'autres niveaux du traitement psycholinguistique (syntaxique ou sémantique) (Taylor & Wales, 1985). Il semble en revanche que le signalement des contours de continuité, outre les variations de la fréquence fondamentale et de l'amplitude, repose sur l'exploitation de tout ou partie du silence subséquent.

Le maniement de l'outil de segmentation tel qu'il a été examiné ici semble apte à accomplir une double mission. Il peut permettre au chercheur de réaliser rapidement un premier examen perceptif de corpus authentiques en vue d'y reconnaître les unités prosodiques participant à l'organisation de ce type de parole. Ceci quel que soit le volume du matériel à analyser. Employé dans le cadre d'études comportementales conçues spécifiquement pour cela, il peut également permettre, d'évaluer la pertinence de la présence ou de la variation de certaines caractéristiques acoustico-prosodiques et de déterminer quels sont les indices que l'auditeur utilise effectivement. Comme par exemple de déterminer dans quels cas les pauses silencieuses sont nécessaires à la perception des frontières.

REFERENCES

- ASTESANO, C., BESSON, M. & ALTER, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, 18, 172-184.
- AURAN, C., COLAS, A., PORTES, C., & VION, M. (2005). Perception of breaks and discourse boundaries in spontaneous speech: Developing an on-line technique. Paper presented at the *IDP05 International Symposium - Discourse and Prosodic Interface Complex*, September 8-9, Aix-en-Provence, France.
- AURAN, C., PORTES, C., RAMI, E. & RIGAUD, N. (2001). La distinction entre frontières conclusives et continuatives est-elle pertinente dans le discours spontané ? *Actes des Journées Prosodie, 10-11 Octobre, 2001, Grenoble, France*, p. 75-78.
- BOERSMA, P. & WEENINK, D. (2005). Praat. Doing phonetics by computer [computer program]. Available from <http://www.praat.org>.
- CHO, H. & HIRST, D. (2006). The contribution of silent pauses to the perception of prosodic boundaries in Lorean read speech. Paper presented at the *Speech Prosody International Symposium*, Dresden, Germany.
- COHEN, H., DOUAIRE, J. & ELSABBAGH, M. (2001). The role of prosody in Discourse processing. *Brain and Cognition* (special issue TENNET), 46 (1-2), 73-82.
- COLAS, A. (1999). Introducing infants to referential events: A developmental study of maternal ostensive marking in French. *Journal of Child Language*, 26, 113-131.
- CUTLER, A., DAHAN, D., & DONSELAAR, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40 (2), 141-201.
- DANIELI, M., GARRIDO, J.-M., MONEGLIA, M., PANIZZA, A., QUAZZA, S., & SWERTS, M. (2004). Evaluation of consensus on the annotation of prosodic breaks in the Romance corpus of spontaneous speech C-ORAL-ROM. Paper presented at the *LREC Conference*, ELRA, Paris.
- DELATTRE, P. (1966). Les dix intonations de base du français. *The French review*, XL, n°1, 1-14.

- DI CRISTO, A. (2004). La prosodie au carrefour de la phonétique, de la phonologie et de l'articulation formes-fonctions. *Travaux Interdisciplinaires Par le et Langage*, 23, 67-211.
- DUTOIT, T., PAGEL, V., PIERRET, N., BATAILLE, F. & VAN DER VREKEN, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. *Proceedings of ICSLP'96*, Philadelphia, USA, Vol. 3, p. 1393-1396.
- ESSER, J. (2000). Medium-transferability and presentation structure in speech and writing. *Journal of Pragmatics*, 32, 1523-1538.
- FERNALD, A., & SIMON, T. (1984). Expanded intonation contours in mothers' speech to newborn. *Developmental Psychology*, 20 (1), 104-113.
- FON, Y.-J. J. (2002). *A cross-linguistic study of syntactic and discourse boundary cues in spontaneous speech*. Ohio State University, USA.
- FRAZIER, L., CARLSON, K., & CLIFTON, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Science*, 10 (6), 244-249.
- FRAZIER, L., CLIFTON, C., & CARLSON, K. (2004). Don't break, or do: Prosodic boundary preferences. *Lingua*, 114, 3-27.
- GROSZ, B. & HIRSCHBERG, J. (1992). Some intonational characteristics of discourse structure. In *Proceedings of International Conference on Spoken Language Processing*, Banff, Alberta, Canada, Vol. 1, p. 429-432.
- HIRSCHBERG, J. & NAKATANI, C. H. (1996). A prosodic analysis of discourse segments in direction-giving monologues. *Proceedings of the Thirty-fourth Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, CA, USA, p. 286-293.
- HRUSKA, C., ALTER, K. & STEUBE, A. (2001). Misleading dialogues: Human's brain reaction to prosodic information. Paper presented at the *Oralité et Gestualité International Symposium*, Aix-en-Provence, France.
- LEHISTE, I. (1979). Perception of sentence and paragraph boundaries. In B. Lindblom & S. Ohman, (eds.). *Frontiers of speech communication research*, p. 191-201. Academic Press: New York.
- LINDBLOM, B., & OHMAN, S. (1979). *Frontiers of speech communication research*, p. 191-201. Academic Press: New York.
- NAKATANI, C. H., HIRSCHBERG, J., & GROSZ, B. J. (1995). Discourse structure in Spoken Language: Studies on Speech Corpora. *Proceedings of the AAAI-95 Spring Symposium of empirical Methods in Discourse Interpretation*, Palo Alto, C.A., USA, p. 106-112.
- PASSONNEAU, R. & LITMAN, D. (1993). Intention-based segmentation: Human reliability and correlation with linguistic cues. *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, Columbus, OH, USA, p. 148-155.
- PERRUCHET, P., TYLER, M., GALLAND, N., & PEEREMAN, R. (2004). Learning non adjacent dependencies: no need for algebraic-like computations. *Journal of Experimental Psychology: General*, 133 (4), 573-583.
- PORTES, C. (2000). *Approche du rôle de la prosodie dans la structuration du discours oral en Français*. Note de recherche non-publiée, Université de Provence, Aix-en-Provence, France.

- PORTES, C. (2002). Approche instrumentale et cognitive de la prosodie du discours en français. *Travaux Interdisciplinaires Par le et Langage*, 21, 101-119.
- PORTES, C. (2004). Prosodie et économie du discours : spécificité phonétique, écologie discursive et portée pragmatique de l'intonation d'implication. Doctorat d'Université, Université de Provence, Aix-en-Provence, France.
- RAMUS, F., HAUSER, M., MILLER, C., MORRIS, D., & MEHLER, J. (2000). Language discrimination by human newborn and by cotton-top tamarin monkeys. *Science*, 288, 349-351.
- ROSSI, M. (1999). *L'int`nati`n, le système du Français*. Gap, France : Ophrys.
- SANDERS, L. & NEVILLE, H. (2003). An ERP study of continuous speech processing; Segmentation, semantics, and syntax in native speakers. *C`gnitive Brain Research*, 15, 228-240.
- SILVERMAN, K., BECKMAN, M., PITRELLI, J., OSTENDORF, M., WIGHTMAN, C. M., PRICE, P., PIERREHUMBERT, J. & HIRSCHBERG, J. (1992). ToBI: A standard for labeling English prosody. *Pr`ceedings `f the Sec`nd Internati`nal C`nference `n Sp`ken Language Pr`cessing*, Ban, Vol. 2, p. 867-870.
- STEINHAUER, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, 86 (1), 142-164.
- STEINHAUER, K., & FRIEDERICI, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *J`urnal `f Psych`linguistic Research*, 30 (3), 267-295.
- STEINHAUER, K., ALTER, K., & FIEDERICI, A.D., (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neur`science*, 2, 191-196.
- STRANGERT, E. (2004). Speech chunks in conversation: Syntactic and prosodic aspects. *Pr`ceedings `f Speech Pr`s`dy*, Nara, Japan, p. 305-308.
- STRANGERT, E. & HELDNER, M. (1995). Labelling of boundaries and prominences by phonetically experienced and non-experienced transcribers. *PHONUM*, 3, 85-109.
- SWERTS, M. (1997). Prosodic features at discourse boundaries of different strength. *J`urnal `f the Ac`ustical S`ciety `f America*, 101, 514-521.
- SWERTS, M. & GELUYKENS, R. (1993). The prosody of information units in spontaneous monologue. *Ph`netica*, 50, 189-196.
- SWERTS, M. & GELUYKENS, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, 37, 21-43.
- TAYLOR, S. & WALES, R. (1987). Primitive mechanisms of accent perception. *J`urnal `f Ph`netics*, 15, 235-246.
- VION, M. & COLAS, A. (2005). La planification des unités prosodiques dans la narration : contrôle intentionnel et contraintes opérationnelles. *Travaux Interdisciplinaires Par le et Langage*, 24, 181-196.
- VION, M. & COLAS, A. (soumis). Intentional control and operational constraints in prosodic phrasing: A study of picture-elicited narations from French children. *Language and Speech*.

ANNEXE : Analyses auditive de l'échantillon expérimental.

Légende : analyse de la version originale (**O**) par les experts (lignes a et b) et par les auditeurs tout-venant (ligne c, réponses données par au moins 6 participants sur 15) ; analyse des versions filtrée (**F**, ligne d), re-synthétisée (**R**, ligne e) et saltra~dzajn (**S**, ligne f) par les auditeurs tout-venant (réponses données par au moins 6 participants).

(a) Contours intonatifs	O									
	O				c. min					c. min
(b) Jointure entre mots										
			<i>non</i>	<i>je</i>	<i>crois</i>	<i>qu'effectiv(e)ment</i>	<i>euh</i>	<i>il</i>	<i>aurait</i>	<i>fallu</i>
(c) Originale	O		non	je	crois	qu'effectiv(e)men /ent	euh	il	aurait	fallu
(d) Filtrée	F		non	je	crois	qu'effectiv(e)ment	eu /euh	il	aurait	fallu
(e) Re-synthétisée	R	a~ /	non	je	crois	qu'effectiv(e)men /ent	euh	il	aurait /	fallu
(f) Saltra~dzajn	S	a~ /	na~	za	tRja	tasattazna~ /	a	al	aRa	sala /

		c. min		c. min				c. maj					
		<i>pouvoir</i>	<i>euh</i>	<i>modifier</i>	<i>les</i>	<i>institutions</i>	<i>européennes</i>	<i>lorsque</i>	<i>nous</i>	<i>sommes</i>	<i>passés</i>		
O		pouvoir	eu / euh	modifier	les	institutions	européennes /	lorsque	nous	sommes	passés		
F		pou / voir	eu / euh	modifier	les	institutions	européennes	lorsque	nous	sommes	passés		
R		pouvoir	eu / euh	modifier	les	institutions	européennes	lorsque	nous	sommes	passés		
S		tazjaRr	a / a	nadasja	la	za~stata	aRataan	laRsta	na	san	tasa		

					implication					c. min		
		<i>de</i>	<i>douze</i>	<i>à</i>	<i>quinze</i>	<i>euh</i>	<i>sur</i>	<i>le</i>	<i>fait</i>	<i>que</i>	<i>l'Autriche</i>	
O		de	douze	à	quinze /	euh /	sur	le	fait	que /	l'Autriche	
F		de	douze	à	quinze / e	euh /	sur	le	fait	que /	l'Autriche	
R		de	douze	à	quin / ze	euh	sur	le	fait	qu / e	l'Autriche	
S		da	daz	a	ta~z / za /	a /	saR	la	sat	ta /	latRas	

				c. min		c. min			c. maj		c. min
		<i>et les</i>	<i>pays</i>	<i>scandinaves</i>	<i>avaient</i>	<i>vocation</i>	<i>à</i>	<i>rejoindre</i>	<i>l'Union</i>	<i>Européenne</i>	<i>personne</i>
O		et les	pays	scandinav / ves	avaient	vocation	à	rejoindre	l'Union	Européenne /	personne
F		et les	pays	scandinaves	avaient	vocation	à	rejoindre	l'Union	Européenne //	personne
R		et les	pays	scandinav / ves	avaient	vocation	à	rejoindre	l'Union	Européenne /	personne
S		ala	ta	sta~danaz //	aza	zatasja~	a	Razja~dR	lanja~	aRataan /	taRsan

	c. min			implication				c. min	
	<i>ne</i>	<i>pouvait</i>	<i>le mettre en en</i>	<i>question</i>	<i>de</i>	<i>la</i>	<i>même</i>	<i>manière</i>	<i>que</i>
O	ne	pouvait	le mettre en en	question	de	la	même	manière	que
F	ne	pouvait	le mettre en en	question	de	la	même	manière	que
R	ne	pouvait	le mettre en en	question	de	la	même	manière	que
S	na	taza	InatRa~/ a~	tastja~	d	la	nan	nanjaR	ta

	c. min				implication			c. min	
	<i>pour</i>	<i>la</i>	<i>Pologne</i>	<i>ou</i>	<i>la</i>	<i>République</i>	<i>tchèque</i>	<i>la</i>	<i>question</i>
O	pour	la	Pologne	ou	la	République	tchèque /	la	question
F	pour	la	Pologne	ou	la	République	tchèque / /	la	question
R	pour	la	Pologne	ou	la	République	tchè/que	la	question
S	taR	la	talanj	a	la	Ratadlat	tsat /	la	tastja~

	c. maj									
	<i>qu'il</i>	<i>fallait</i>	<i>se poser</i>	<i>qu'on</i>	<i>s'est</i>	<i>posée</i>	<i>c'est de</i>	<i>savoir</i>	<i>si</i>	<i>les</i>
O	qu'il	fallait	se poser	qu'on	s'est	posée	c'est de	savoir	si	les
F	qu'il	fallait	se poser	qu'on	s'est	posée	c'est de	savoir	si	/es
R	qu'il	fallait	se poser	qu'on	s'est	posée	c'est de	savoir	si	les
S	ta	sala	staza	ta~	sa	taza	d	sazjaR/	sa	la

	c. maj			c. maj			c. min		
	<i>institutions</i>	<i>européennes</i>	<i>étaient</i>	<i>capables</i>	<i>de</i>	<i>fonctionner</i>	<i>aussi</i>	<i>bien</i>	<i>à quinze</i>
O	institutions	européennes/	étaient	capables	de	fonctionner	aussi	bien	à quinze
F	institutions	européennes/	étaient	capables	de	fonctionner	aussi	bien	à quinze
R	institutions	européennes	étaient	capables	de	fonctionner	aussi	bien	à quinze
S	za~statajsa~	aRataan /	ata	tatadla	da	s/a~tsjana	asa	dja~	ata~z

	c. maj					finalité			
	<i>vingt</i>	<i>ou</i>	<i>vingt-cinq</i>	<i>qu'elles</i>	<i>fonctionnaient</i>	<i>à</i>	<i>douze</i>	<i>ça</i>	<i>n'était pas</i>
O	vingt	ou	vingt-cinq	qu'elles	fonctionnaient	à	douze	ça	n'était pas
F	vingt	ou	vingt-cinq	qu'elles	fonctionnaient	à	douze	ça	n'était pas
R	vingt	ou	vingt-cinq	qu'elles	fonctionnaient	à	douze	sa	n'était pas
S	za~	a	za~tsa~	tal	sa~tsjana	a	daz	sa	nata ta

		c.maj				c.maj			
	<i>le</i>	cas	<i>euh</i>	<i>à</i>	<i>l'époque</i>	<i>euh</i>	<i>le</i>	<i>président</i>	<i>de</i>
O	le	cas /	euh	à	l'époque	eu / euh	le	président	de
F	le	cas /	euh	à	l'époque	euh	le	président	de
R	le	cas /	euh	à	l'époque	euh	le	président	de
S	la	ta /	a	a	latata	a	la	tRazada~	d

		c.min				c.maj		c.maj		
	<i>la</i>	<i>République</i>	<i>qui</i>	<i>était</i>	<i>le</i>	<i>Président</i>	Mitterrand	a choisi	<i>euh</i>	<i>de</i>
O	la	République	qui	était	le	Président	Mitterrand /	a choisi /	euh	de
F	la	République	qui	était	le	Président	Mitterrand	a / choisi	euh	de
R	la	République	qui	était	le	Président	Mitterrand	a ch / choisi	euh	de
S	la	Ratadla	ta	ata	la	tRazada~	nataRa~ /	asjaza /	a	da

		c.min					c.maj			
	<i>fai</i>	<i>laisser</i>	<i>entrer</i>	<i>l'Autriche</i>	<i>et</i>	<i>les</i>	<i>pays</i>	<i>scandinaves</i>	<i>avant</i>	<i>de</i>
O	fai	laisser	entrer	l'Autriche	et	les	pays	scandina / aves	avant	de
F	fai	laisser	entrer	l'Autriche	et	les	pays	scandinave / es	a / vant	de
R	fai	laisser	entrer	l'Autriche	et	les	pays	scandinav / ves	avant	de
S	sa	lasa	a~tRa	latRas	a	la	taa	sta~danaz /	aza~	da

		c.min		c.maj			c.min			
	<i>faire</i>	<i>la</i>	<i>réforme</i>	<i>institutionnelle</i>	<i>qui</i>	<i>a</i>	<i>été</i>	<i>renvoyée</i>	<i>à</i>	<i>l'élargiss(e)ment</i>
O	faire	la	réforme	institutionnelle /	qui	a	été	renvoyée	à	l'élargiss(e)ment
F	faire	la	réforme	institutionnelle /	qui	a	été	renvoyée	à	l'élargiss(e)ment
R	faire	la /	réforme	institutionnelle	qui	a	été	renvoyée	à	l'élargiss(e)ment
S	saR	la	RasaRn	a~statajanal /	ta	a	ata	Ra~zjaja	a	lalaRzasna~

		finalité				c.min		c.maj		
	suivant	<i>et</i>	<i>nous</i>	<i>nous</i>	<i>sommes</i>	<i>retrouvés</i>	<i>au</i>	<i>traité</i>	d'Amsterdam	<i>qui</i>
O	suivant / /	et	nous	nous	sommes	retrouvés	au	traité	d'Amsterdam	qui
F	suivant	et	nous	nous	sommes	retrouvés	au	traité	d'Amsterdam	qui
R	suivant /	et	nous	nous	sommes	retrouvés	au	traité	d'Amsterdam	qui
S	sjaza~ /	a	na	na	san	RatRaza	a	tRata	danstaRdan /	ta

				c. min	c. maj		c. min		
	<i>n'est</i>	<i>pas</i>	<i>non</i>	<i>plus</i>	satisfaisant	<i>et</i>	<i>donc</i>	<i>il</i>	<i>faut</i>
O	n'est	pas	non	plus	satisfaisant /	et	donc	il	faut
F	n'est	pas	non	plus	satisfaisant	et	donc	il	faut
R	n'est	pas	non	plus	satisfaisant	et	donc	il	faut
S	na	ta	na~	tla	satassaza~ /	a	da~ ta	al	sa

	c. min				finalité
	aujourd'hui remettre	l'ouvrage	sur	le	métier
O	aujourd'hui remettre	l'ouvrage	sur	le	métier /
F	aujourd'hui remettre	l'ouvrage	sur	le	métier
R	aujourd'hui remettre	l'ouvrage	sur	le	métier /
S	azaRdja	RanatR	lazRaz	saR	la natja